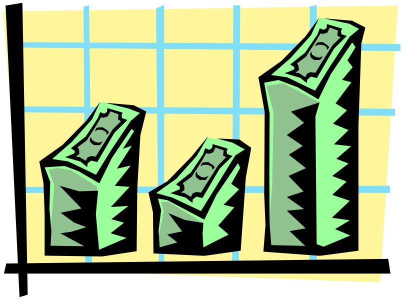
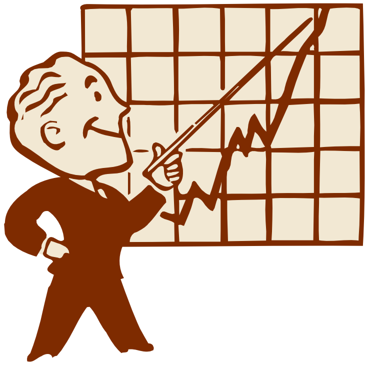




Introduction to Probability and Statistics

Chapter 7

Ammar M. Sarhan,
asarhan@mathstat.dal.ca
Department of Mathematics and Statistics,
Dalhousie University
Fall Semester 2008

Chapter 7

Statistical Intervals Based on a Single Sample

Confidence Intervals

An alternative to reporting a single value for the parameter being estimated is to calculate and report an entire interval of plausible values – a *confidence interval (CI)*. *A confidence level* is a measure of the degree of reliability of the interval.

7.1 Basic Properties of Confidence Intervals

If after observing $X_1 = x_1, \ldots, X_n = x_n$, *we* compute the observed sample mean $\bar{x}$, then a *95%* confidence interval for μ can be expressed as

$$\left(\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right)$$

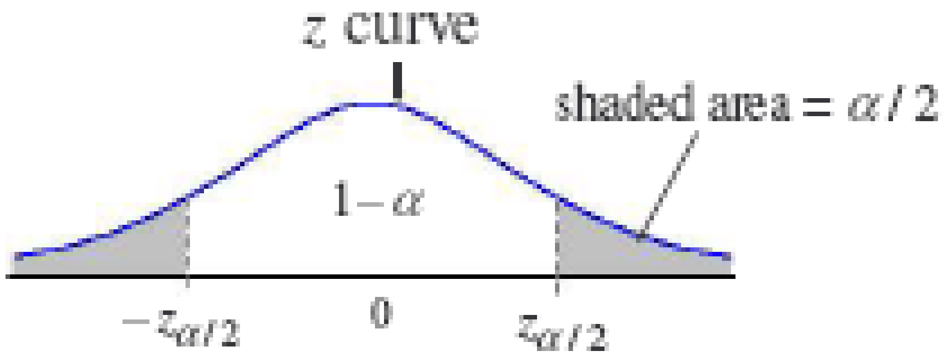
The population has a $N(\mu, \sigma^2)$ and σ is known.

Other Levels of Confidence

A (1- α)100% confidence interval for the mean μ of a normal population when the value of is known α is given by

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Other Levels of Confidence



$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1-\alpha$$

Sample Size

The general formula for the sample size n necessary to ensure an interval width w is

$$n = \left(2 z_{\alpha/2} \cdot \frac{\sigma}{w}\right)^2$$

Deriving a Confidence Interval

Let $X_1, \ldots, X_n$ denote the sample on which the CI for the parameter θ is to be based.

Suppose a ***random variable*** satisfying the following properties can be found:

1. The variable depends functionally on both $X_1, \ldots, X_n$ and θ.
2. The probability distribution of the variable ***does not depend*** on θ or any other unknown parameters.

Let $h(X_1, \ldots, X_n; \theta)$ denote this random variable.

In general, the form of h is usually suggested by examining the distribution of an appropriate estimator $\hat{\theta}$.

For any α, $0 < \alpha < 1$, constants a and b can be found to satisfy

$$P\left(a < h(X_1, \cdots, X_n; \theta) < b\right) = 1 - \alpha$$

Now suppose that the inequalities can be manipulated to isolate θ

$$P\left(l(X_1, \cdots, X_n) < \theta < u(X_1, \cdots, X_n)\right) = 1 - \alpha$$

lower confidence limit

upper confidence limit

For a 100(1- α)% CI.

Examples: 7.5 p. 260. *will be given in the class.*

Large-Sample Confidence Interval

Let $X_1, \ldots, X_n$ denote the sample from a population having a mean μ and standard deviation σ. If n is sufficiently large, then

$$\bar{X} \sim N\left(\mu, \tfrac{\sigma^2}{n}\right) \quad \Longrightarrow \quad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

This implies that

$$\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

is **a large-sample confidence interval for** **μ** with level 100(1- α)%.

This formula is valid regardless of the shape of the population distribution.

For practice: $n > 40$.

Notice, if σ is unknown, replace it with the sample standard deviation s.
That is,

$$\left(\overline{X} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \overline{X} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \right)$$

Examples: 7.6 p. 264.
Suppose a random sample with size 48 from a population with unknown mean μ and unknown variance σ^2 with the following information:

$$\sum x_i = 2626 \text{ , and } \sum x_i^2 = 144,950$$

Find the 95% confidence interval of μ?

Solution: From the information given in the problem, we have:

$$\overline{x} = 54.7 \text{ and } s = 5.23$$

Then, the 95% confidence interval of μ is

$$\left(54.7 - 1.96 \frac{5.23}{\sqrt{48}}, 54.7 + 1.96 \frac{5.23}{\sqrt{48}} \right) = (53.2, 56.2)$$

That is, with a confidence level of approximation 95%,

$$53.2 < \mu < 56.2$$

Confidence Interval for a Population Proportion

Let p denote the proportion of "*successes*" in a population, where *success* identifies an individual or object that has a specified property.
A random sample of n individuals is to be selected, and X is *the number of successes in the sample*.
A confidence interval for a *population proportion p with level* 100(1- α)% is:

$$\left(\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \left(z_{\alpha/2}\right)^2 / n}, \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \left(z_{\alpha/2}\right)^2 / n} \right)$$

where, $\hat{p} = \dfrac{X}{n}$, $\hat{q} = 1 - \hat{p}$.

Notice, if the sample size is quite large, the approximate CI limits become

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}} \right)$$

Since $z^2_{\alpha/2}/(2n)$ is negligible compared to $\hat{p}$.

Sample Size

The general formula for the sample size n necessary to ensure an interval width w is

$$n \approx \frac{4\, z^2_{\alpha/2}\, \hat{p}\,\hat{q}}{w^2}$$

Examples: 7.8 p. 267.

Suppose that in 48 trials in a particular laboratory, 16 resulted in ignition of a particular type of substrate by a lighted cigarette.

Find the 95% confidence interval of the long-run proportion of all such trails that would result in ignition?

Solution: $n = 48$

Let p denote the long-run proportion of all such trails that would result in ignition.

$$\hat{p} = \frac{X}{n} = \frac{16}{48} = \frac{1}{3} = 0.333, \quad q = 0.667$$

Then, the approximately 95% CI of p is

$$\frac{0.333 + \frac{(1.96)^2}{2(48)} \pm 1.96\sqrt{\frac{(0.333)(0.667)}{48} + \frac{(1.96)^2}{4(48)^2}}}{1 + (1.96)^2 / 48}$$

$$= \frac{0.333 \pm 0.139}{1.08} = (0.217, 0.474)$$

Notice, the traditional 95%CI is

$$0.333 \pm 1.96\sqrt{\frac{(0.333)(0.667)}{48}} = (0.200, 0.466)$$

Examples: 7.9 p. 267.

Find the sample size necessary to ensure a width of 0.10 for the 95% confidence interval of the long-run proportion of all such trails that would result in ignition?

$$n \approx \frac{4 z_{\alpha/2}^{2} \hat{p} \hat{q}}{w^{2}}$$

$$=4*(1.96)^{2} * 0.333*0.667 / .01 = 341.305$$

n =341

7.3 Intervals Based on a Normal Population Distribution

The population of interest is normal, so that $X_1, \ldots, X_n$ constitutes a random sample from a *normal distribution* with both μ and σ^2 unknown.

t Distribution

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, then the rv

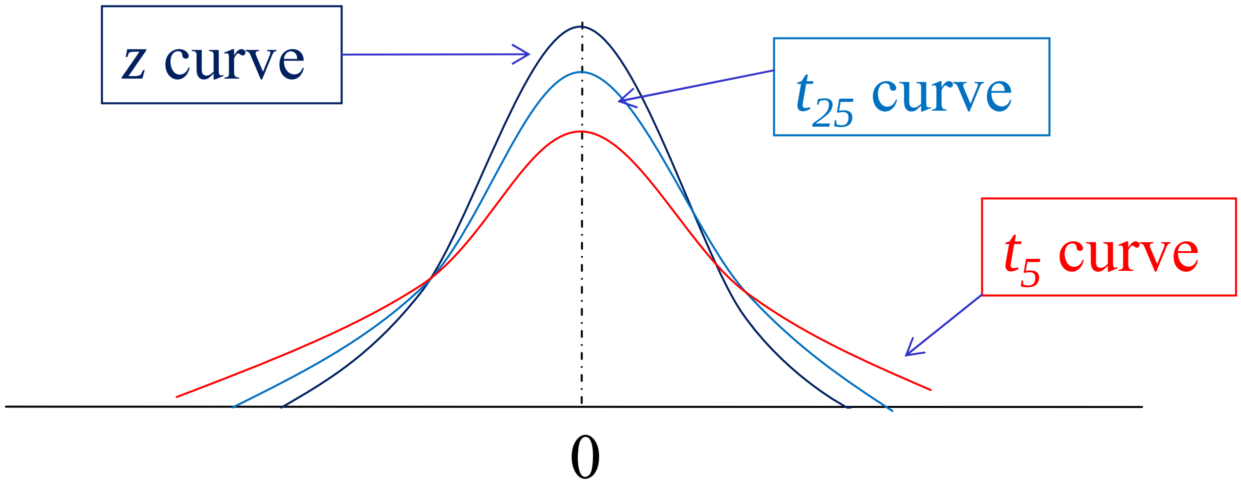
$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

has a probability distribution called a ***t distribution*** with n-1 degrees of freedom (df).

Properties of *t Distributions*

Let t_v denote the density function curve for v df.

1. Each t_v curve is bell-shaped and centered at 0.
2. Each t_v curve is spread out more than the standard normal (z) curve.
3. As v increases, the spread of the corresponding t_v curve decreases.
4. As $v \to \infty$, the sequence of t_v curves approaches the standard normal curve (the z curve is called *a t* curve with df $= \infty$)



t Critical Value

Let $t_{\alpha,v}$ = the number on the measurement axis for which the area under the t curve with v df to the right of $t_{\alpha,v}$ is α.
$t_{\alpha,v}$ is called *t* critical value.

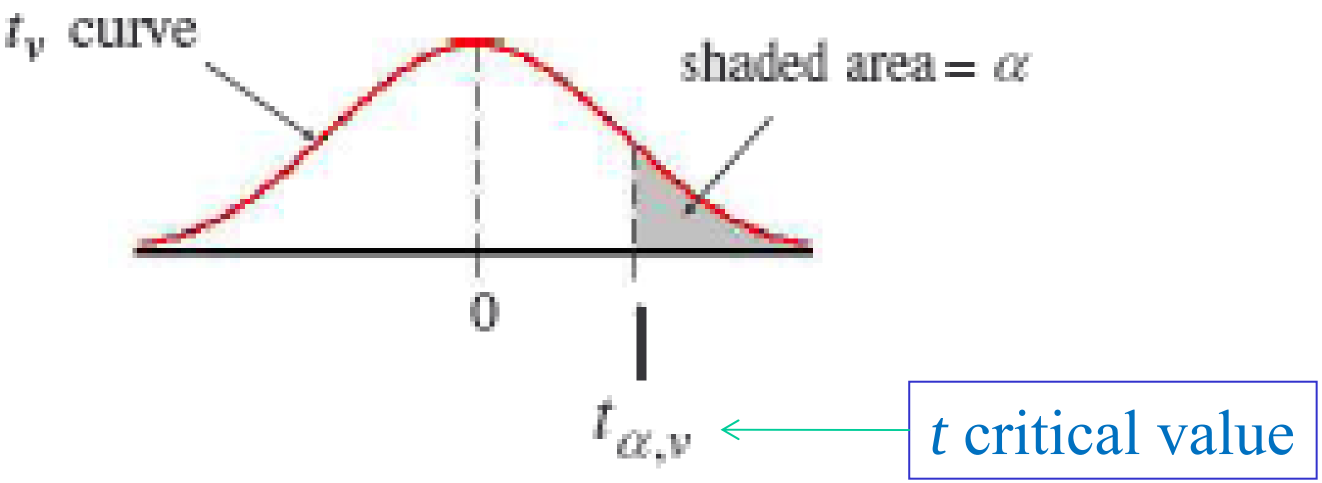


Table A.5, p. 671, gives the t critical value for given α, *v*.

$t_{0.025,15}$ =2.131, $t_{0.05,22}$ =1.717 , $t_{0.01,22}$ =2.508.

t Confidence Interval

Now, let $X_1,\ldots, X_n$ be a random sample from a *normal distribution* with both μ and σ^2 unknown, then the $(1-\alpha)100\%$ CI of μ is

$$\left(\overline{X} - t_{\alpha/2,n-1}\cdot\frac{S}{\sqrt{n}},\ \overline{X} + t_{\alpha/2,n-1}\cdot\frac{S}{\sqrt{n}}\right)$$

or

$$\overline{X} \pm t_{\alpha/2,n-1}\cdot\frac{s}{\sqrt{n}}$$

Here, $\overline{X}$ and S are the sample mean and sample variance.

Examples: 7.12 p. 274.

Consider the following sample of fat content (in percentage) on 10 randomly selected hot dogs:

25.2 21.3 22.8 17.0 29.8 21.0 25.5 16.0 20.9 19.5

Find the 95% confidence interval of the population mean fat content, assuming the population is normal.?

Solution: We have:

$$n = 10, \quad \bar{x} = 21.90 \quad \text{and} \quad s = 4.134, \quad \alpha/2 = 0.025$$

Then, the 95% confidence interval of μ is

$$\left(21.90 - t_{0.025,10-1} \cdot \frac{4.134}{10}, 21.90 + t_{0.025,10-1} \cdot \frac{4.134}{10} \right)$$

$$= \left(21.90 - 2.262 \cdot \frac{4.134}{10}, 21.90 + 2.262 \cdot \frac{4.134}{10} \right)$$

$$= (18.94, 24.86)$$

Summary

The random sample	$(1-\alpha)100\%$ CI of μ
$X_1,\ldots, X_n \sim N(\mu, \sigma^2)$ σ is known	$\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
$X_1,\ldots, X_n \sim N(\mu, \sigma^2)$ σ is unknown	$\bar{X} \pm t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}$
$X_1,\ldots, X_n \sim$ any population with mean μ and variance σ^2 σ is known ($n \geq 30$)	$\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
σ is unknown ($n \geq 30$)	$\bar{X} \pm z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$